DOCUMENT RESUME

ED 459 202 TM 033 499

AUTHOR Newman, Isadore; Fraas, John W.; Herbert, Alan
TITLE Testing Non-Nil Null Hypotheses with t Tests of Group Means: A Monte Carlo Study.
PUB DATE 2001-10-00
NOTE 29p.; Paper presented at the Annual Meeting of the Mid-Western Educational Research Association (Chicago, IL, October 24-27, 2001).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Groups; *Hypothesis Testing; Monte Carlo Methods; *Statistical Significance
IDENTIFIERS *Mean (Statistics); *Null Hypothesis; T Test

ABSTRACT

Statistical significance and practical significance can be considered jointly through the use of non-nil null hypotheses that are based on values deemed to be practically significant. When examining differences between the means of two groups, researchers can use a randomization test or an independent t test. The issue addressed in this paper is whether Type I error rates produced by independent t tests of group means are impacted by the use of non-nil null hypotheses. The results of this Monte Carlo study suggest the frequencies of Type I error rates produced by independent t tests of group means for non-nil hypotheses are comparable to those recorded for nil null hypotheses. These results should provide further encouragement for researchers to use non-nil hypotheses as a means of considering statistical and practical significance jointly. (Contains 5 tables and 15 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 459 202

Running head: TESTING NON-NIL NULL HYPOTHESES

Testing Non-Nil Null Hypotheses with t Tests of Group Means: A Monte Carlo Study

Isadore Newman
The University of Akron

John W. Fraas
Ashland University

Alan Herbert
The University of Akron

Paper presented at the annual meeting of the
Mid-Western Educational Research Association
Chicago, Illinois
October 24-27, 2001

TM033499

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
J. W. Fraas
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

BEST COPY AVAILABLE

Abstract

Statistical significance and practical significance can jointly be considered through the use of non-nil null hypotheses which are based on values deemed to be practically significant. When examining differences between the means of two groups, researchers can utilize a randomization test or an independent t test. The issue addressed in this paper is whether Type I error rates produced by independent t tests of group means are impacted by the use of non-nil null hypotheses. The results of this Monte Carlo study suggest the frequencies of Type I error rates produced by independent t tests of group means for non-nil null hypotheses are comparable to those recorded for nil null hypotheses. These results should provide further encouragement for researchers to utilize non-nil null hypotheses as a means of jointly considering statistical and practical significance.

Testing Non-Nil Null Hypotheses with t Tests of Group Means: A Monte Carlo Study

Researchers can choose to use nil null hypotheses or non-nil null hypotheses. A nil null hypothesis is based on the proposition that the two group means do not differ; while a non-nil null hypothesis is based on the assumption that the two group means do not differ by more than some specified value. Thus, for a nil null hypothesis any difference greater than zero is deemed to be important by the researchers. For a non-nil null hypothesis, however, the researchers must determine the size of difference between the two group means that appears to be relevant to the factors being studied.

Stressing the need for researchers and practitioners to identify the size of the difference between the group means that is deemed important, i.e., practically significant, reflects the recent stress placed on considering practical significance along with statistical significance. Cohen (1988, 1994), Huberty (1993), Robinson and Levin (1997), Shaver (1993), and Thompson (1996, 1999a, 1999b, 1999c), strongly suggest current research practice should incorporate the reporting and interpreting of measures of practical significance. Robinson and Levin argued the results of statistical hypothesis testing should be conducted and reported along with the effect sizes. Cohen (1994) even recommended researchers should use non-nil null hypothesis. Cohen stated: "Even null hypothesis testing complete with power analysis can be useful if we abandon the rejection of the point nil hypotheses" (p. 1002).

Building on the position stated by Cohen (1994), Fraas and Newman (2000, 2001) suggested current research practices would be strengthened if researchers incorporated practical significance levels into the construction and statistical testing of their null hypotheses. That is, a

researcher should use a non-nil null hypothesis that states the difference between two group means is equal to or less than some value other than zero. Fraas and Newman proposed this value be set equal to a value researchers and practitioners believe the difference between the two group means must exceed in order for the difference to be practically or clinically significant.

If the researchers are truly interested in differences between group means that only exceed a given value, it is essential to incorporate the specified value into the hypothesis testing procedure. If such a practice is not followed, the researchers are not matching the statistical procedure to the research question. Newman, Deitchman, Burkholder, Sanders, and Ervin (1976) identified this type of inconsistency between the statistical procedure employed by the researchers and the research question as a Type VI error.

Fraas and Newman (2000) suggested one important question should be addressed regarding the use of non-nil null hypotheses is: How can the current major statistical computer software packages be used in conjunction with the testing of non-nil null hypotheses? This question is critical to investigate due to the likelihood that unless researchers are able to test non-nil null hypotheses with readily available computer software, they may continue to exclusively use nil null hypotheses. With respect to using readily available computer software in conjunction with non-nil null hypotheses, Thompson (1999a) noted most computer packages assume the researchers are testing nil null hypotheses. Thus, they are not equipped to invoke the necessary changes in calculations. Selin and Lapsley (1985, 1993) suggested such changes include the use of critical values obtained from noncentralized t and F distributions. In addition, Thompson stated some of the complexities of using non-nil null hypotheses are not yet readily applicable in many designs.

Fraas and Newman (2000, 2001) expressed the view that if techniques used to test non-nil null hypotheses are not readily available to researchers, they will tend not to use such hypotheses. Since researchers currently employ t tests and such tests are readily available in current computer software, the question addressed in this study is: Can a non-nil null hypothesis designed to test the difference between two group means at a given practical significance level be conducted with a t test without substantially affecting its type I error rate?

To investigate this question we compared, by means of a Monte Carlo study, Type I error rates produced by independent t tests used to statistically test non-nil null hypotheses and nil-null hypotheses. Specifically, this study was designed to investigate the impact various combinations of the following factors have on the relative number of Type I error rates: (a) normally distributed populations, (b) non-normally distributed populations, (c) equal and unequal sample sizes, (d) equal and unequal population variances, and (e) the size of the values incorporated into the non-nil null hypotheses. Comparable Type I error rate frequencies would suggest an independent t test of the differences between group means is robust with respect to the testing of non-nil null hypotheses. Such a result, which would provide researchers with more latitude in selecting a testing technique for non-nil null hypotheses, will eliminate a possible barrier to the use of non-nil null hypotheses which incorporate practically significant values.

Parameters of the Monte Carlo Study

The goal of this Monte Carlo study was to compare the numbers of Type I error rates per test produced by independent t tests used to test nil null hypotheses and non-nil null hypotheses under various population parameters and sample sizes. One set of results was produced for normally distributed populations generated by the normal distribution random number generator

contained in the Microsoft Excel Professional 2000 (2000) computer software. Another set of results was obtained for non-normally distributed populations generated by the Poisson distribution generator also contained in the Microsoft Excel Professional 2000 computer software. Normally Distributed Populations

Twelve normally distributed populations were generated. Two of the 12 populations, labeled C1 and C2, were identified as control populations. The other 10 populations were labeled as experimental populations (E1 to E12). Two of the 12 experimental populations, labeled E1 and E6, were generated and used in conjunction with the two control populations to estimate the number of Type I error rates produced by the independent t tests of group means under the conditions stated in nil null hypotheses. Each nil null hypothesis, along with the corresponding research hypothesis, was designed to determine whether the mean of a given experimental group was statistically significantly higher than the mean of a specified control group. It is important to note each nil null hypothesis tested was based on the assumption that any value by which the mean of the experimental group exceeded the mean of the control group was deemed to be important to test by the researchers. Thus, to conduct a Monte Carlo study on the number of Type I error rates per test produced in conjunction with the nil null hypotheses under this assumption, the means of E1 and E6 ($\bar{x} = 25.0$) were set equal to the means of the C1 and C2 ($\bar{x} = 25.0$). Since one factor allowed to vary in this study was the population variances, the variances of C1 and E1 were set at 2.0; while the variances of C2 and E6 were set at 4.0.

The means of the other eight experimental populations were set at levels that allowed the number of Type I error rates to be estimated for independent t tests of group means under the conditions stated in non-nil null hypotheses. That is, the mean of each of these eight

experimental populations was set a level that would exceed the mean of the control population by an amount equal to the practical significance level incorporated into the corresponding non-nil null hypothesis. In this study, the following four practical significance levels were incorporated into the non-nil null hypotheses: (a) .3 of a point, (b) .7 of a point, (c) 1.1 points, and (d) 5.0 points. It may be helpful to relate these practical significance levels to effect sizes. For the purpose of this Monte Carlo study, an effect size represented the difference between the mean of the experimental population and the mean of the control population divided by the smaller standard deviation value of the two populations. Thus, the practical significance levels of .3, .7, 1.1, and 5.0 reflected differences between group means that, according to Cohen (1989), would be classified as small, medium, large, and very large effect sizes, respectively.

Based on the selected practical significance levels, eight of the experimental populations, labeled E2 through E5 and E7 through E10, were generated with means set at levels .3, .7, 1.1, and 5.0 points higher than the means of the control groups. Specifically, the means of these eight experimental populations were as follows: (a) 25.3 for E2 and E7, (b) 25.7 for E3 and E8, (c) 26.1 for E4 and E9, and (d) 30.0 E5 and E10.

A second factor allowed to vary in this Mont Carlo study was the size of the adjustment in the non-nil null hypothesis. Various adjustment sizes were generated by comparing the group means randomly selected from the following populations:

1. Comparison of C1 to E2 and C2 to E7 – small effect size adjustments.
2. Comparison of C1 to E3 and C2 to E8 – moderate effect size adjustments.
3. Comparison of C1 to E4 and C2 to E9 – large effect size adjustments.
4. Comparison of C1 to E5 and C2 to E10 – very large effect size adjustments.

As previously stated, this study was undertaken to investigate the impact of various combinations of variance values. Thus, the variances of these eight experimental populations were set at two different levels. Populations E2 through E5 had variances of 2.0; while the populations E7 through E10 had variances of 4.0. The parameters of the two control populations and the 10 experimental populations are listed in Table 1.

Insert Table 1 about here

Calculation of Type I error rates for the normally distributed populations. The numbers of Type I error rates per test produced by independent t tests conducted on the differences between group means were tabulated for the nil null hypotheses and the non-nil null hypotheses. To obtain Type I error rates per test for nil null hypotheses, the following three nil null hypotheses were constructed:

1. Nil Null Hypothesis 1: H_0: $\mu_{C1} = \mu_{E1}$
2. Nil Null Hypothesis 2: H_0: $\mu_{C1} = \mu_{E6}$
3. Nil Null Hypothesis 3: H_0: $\mu_{C2} = \mu_{E1}$.

The populations included in these nil null hypotheses reflected various combinations of variance parameters. Nil Null Hypothesis1 was tested under the parameter conditions where the variances of C1 and E1 were equal at a level of 2.0. Nil Null Hypothesis 2 was tested under the parameter conditions where the variances of C1 and E6 were unequal. The variance of C1 was 2.0; while the variance of E6 was 4.0. The variances of the populations utilized in conjunction with Nil

Null Hypothesis 3 were also unequal. For these populations, the variance of C2 was 4.0; while the variance of E1 was 2.0.

To obtain Type I error rates per test for non-nil null hypotheses, the following twelve non-nil null hypotheses were constructed:

1. Non-Nil Null Hypothesis 1: H_0: $\mu_{C1} = \mu_{E2} - .3$
2. Non-Nil Null Hypothesis 2: H_0: $\mu_{C1} = \mu_{E3} - .7$
3. Non-Nil Null Hypothesis 3: H_0: $\mu_{C1} = \mu_{E4} - 1.1$
4. Non-Nil Null Hypothesis 4: H_0: $\mu_{C1} = \mu_{E5} - 5.0$
5. Non-Nil Null Hypothesis 5: H_0: $\mu_{C1} = \mu_{E7} - .3$
6. Non-Nil Null Hypothesis 6: H_0: $\mu_{C1} = \mu_{E8} - .7$
7. Non-Nil Null Hypothesis 7: H_0: $\mu_{C1} = \mu_{E9} - 1.1$
8. Non-Nil Null Hypothesis 8: H_0: $\mu_{C1} = \mu_{E10} - 5.0$
9. Non-Nil Null Hypothesis 9: H_0: $\mu_{C2} = \mu_{E2} - .3$
10. Non-Nil Null Hypothesis 10: H_0: $\mu_{C2} = \mu_{E3} - .7$
11. Non-Nil Null Hypothesis 11: H_0: $\mu_{C2} = \mu_{E4} - 1.1$
12. Non-Nil Null Hypothesis 12: H_0: $\mu_{C2} = \mu_{E5} - 5.0$.

The populations included in these nil null hypotheses also reflected various combinations of variance values. Non-Nil Null Hypothesis 1 through Non-Nil Hypothesis 4 were tested under the parameter conditions where the variances of the control and experimental populations – C1 and

E2 through E5 – were equal at a level of 2.0. Non-Nil Null Hypothesis 5 through Non-Nil Hypothesis 8 were tested under the parameter conditions where the variances of C1 was 2.0; while the variances of E7 through E10 were 4.0. And Non-Nil Null Hypothesis 9 through Non-Nil Hypothesis 12 were tested under the parameter conditions where the variances of C2 was 4.0; while the variances of E2 through E5 were 2.0.

To estimate the numbers of Type I error rates per test produced by independent t tests of the differences between group means, random samples were selected from a control population and an experimental population. A third factor allowed to vary in this study was group sample size. Thus, each of the three nil null hypotheses and the twelve non-nil null hypotheses were tested under the following three sample size combinations:

1. Random samples of 25 were selected from the control and experimental populations.
2. A random sample of 15 was selected from the control population; while a random sample of 25 was selected from the experimental population.
3. A random sample of 25 was selected from the control population; while a random sample of 15 was selected from the experimental population.

Once the samples were selected from a given combination of control and experimental populations, an independent t test was calculated for the difference between group means; where the mean of the experimental group was subtracted from the mean of the control group. The calculated t value was compared to the one-tailed critical t value at the .05 alpha level. Since the control and experimental populations were equal for each nil null hypothesis, any negative calculated t value less than the negative critical t value was an indication of a Type I error. That

is, a Type I error was committed when the calculated t value was located outside of the critical t value in the t distribution.

To understand how the number of Type I errors were estimated for the non-nil null hypotheses, it is important to know how the t values were calculated for such hypotheses. Before the t test was calculated for a given non-nil null hypothesis, the practical significance value incorporated into the hypothesis was subtracted from the mean of the experimental group. This adjusted experimental group mean was subtracted from the control group mean. Since each experimental population included in a given non-nil null hypothesis was generated in such a manner that its mean differed from the corresponding control population by an amount equal to the practical significance value, a negative calculated t value less than the negative critical t value would indicate a Type I error was committed. That is, a Type I error was committed when the calculated t value was located outside of the critical t value in the t distribution.

The testing procedure for each nil null hypothesis and non-nil null hypotheses was repeated 5000 times. Thus, the number of Type I error rates per test for each null hypothesis was calculated by dividing the total number of Type I errors recorded for the 5000 replications by 5000. A comparison of the numbers of Type I error rates per test produced for the following pairs of hypotheses reveals the relative number of errors recorded for nil null hypotheses and the non-nil null hypotheses:

1. Nil-Null Hypothesis 1 compared to Non-Nil Null Hypotheses 1 through 4.
2. Nil-Null Hypothesis 2 compared to Non-Nil Null Hypotheses 5 through 8.
3. Nil-Null Hypothesis 1 compared to Non-Nil Null Hypotheses 9 through 12.

The numbers of Type I error rates per test produced by the independent t test used in conjunction

with nil null hypotheses and non-nil null hypotheses that involve normally distributed populations are listed in Table 2 through Table 4.

Insert Table 2 through Table 4 about here

Non-Normally Distributed Populations

Three non-normally distributed populations were generated. Each of these three populations reflected a Poisson distribution. One of these three populations, which was labeled C3, was identified as the control population. The other two populations, which were labeled as E11 and E12, were identified as experimental populations. The C3 and the E11 populations were generated to estimate the number of Type I error rates per test produced by the testing of a nil null hypothesis. This nil null hypothesis was designed to statistically test whether the mean of the control group was not less than the mean of an experimental group. As was the case for the nil null hypotheses used in conjunction with the normally distributed populations, the nil null hypothesis constructed for the non-normally distributed populations was based on the assumption that any value by which the mean of the experimental group exceeded the mean of the control group was deemed to be important to test by the researchers. Thus, to conduct a Monte Carlo study on the number of Type I error rates per test produced by the statistical testing of the nil null hypotheses under this assumption, the mean of E11($\bar{x}$ = 25.0) was set equal to the mean of C3 ($\bar{x}$ = 25.0). Since the populations reflected Poisson distributions, the variances of C3 and E11 were equal to their means. That is, the variances of C3 and E11 were 25.0.

The mean of the other experimental population, E12, was set at a level that exceeded the mean of C3 by the practical significance level incorporated into the non-nil null hypothesis. For the non-normally distributed populations, only one practically significant value was investigated. The mean of E12 was set at 30.0, which reflected a difference from the mean of C3 that would be classified as a small effect size. Since E12 reflects a Poisson distribution and it has a mean of 30.0, its variance is 30.0. The parameters of the three non-normally distributed populations – C3, E11, and E12 – are listed in Table 1.

Calculation of Type I Error Rates for the Non-Normally Distributed Populations. The numbers of Type I error rates per test produced by independent t tests conducted in conjunction with nil null hypotheses and non-nil null hypotheses were calculated under a variety of population parameters and sample sizes for the non-normally distributed populations. Type I error rates per test were calculated for one nil null hypothesis. This nil null hypothesis, which was identified as Nil Null Hypothesis 4, was H_0: $\mu_{C3} = \mu_{E11}$. This nil null hypothesis was tested under the parameter condition where the variances of the C3 and E11 were equal at the level 25.0.

Type I error rates per test were obtained for one non-nil null hypothesis. This non-nil null, which was identified as Non-Nil Null Hypothesis 13, was H_0: $\mu_{C3} = \mu_{E12} - 30.0$. Non-Nil Null Hypothesis 13 was tested under the parameter conditions where the variances of C3 and E12 were 25.0 and 30.0, respectively.

To estimate the numbers of Type I error rates per test produced by independent t tests of the differences between group means, random samples were selected from a control population

and an experimental population for each nil null hypothesis and non-nil null hypothesis. Both the nil null hypothesis and the non-nil null hypothesis were tested under the same three sample size combinations used with the normally distributed populations, which were as follows:

4. A random sample of 25 was selected from the control and experimental populations.
2. A random sample of 15 was selected from the control population; while a random sample of 25 was selected from the experimental population.
3. A random sample of 25 was selected from the control population; while a random sample of 15 was selected from the experimental population.

Once the samples were selected, the independent t values were calculated and compared to the critical t values in the same manner as they were for the normally distributed populations, which was presented previously. The testing procedure for the nil null hypothesis and the non-nil null hypothesis was repeated 5000 times. A comparison of the numbers of Type I error rates per test produced for Nil-Null Hypothesis 4 and Non-Nil Null Hypothesis 13 revealed the relative number of errors for the two types of null hypotheses. The number of Type I errors recorded for Nil-Null Hypothesis 4 and Non-Nil Null Hypothesis 13 are listed in Table 5.

Insert Table 5 about here

Results

The number of Type I errors per test were recorded for normally distributed populations in Table 2 through Table 4 and non-normally distributed populations in Table 5. With respect to the normally distributed populations, Tables 2 through Table 4 contain results under three

different combinations of variance values. In Table 2, the Type I error rates per test were produced under conditions where the variance of the control population was equal to the variances of the experimental populations. The results contained in Table 3 were produced under conditions where the experimental populations had variances twice the size of the variance of the control population. And in Table 4, the variance values were reversed. That is, the Type I error rates per test were produced under conditions where the variance of the control population was twice the size of the variances of the experimental populations.

Equal variances for normally distributed populations. In Table 2, the values listed under the column entitled "Combination 1" indicated the number of Type I error rates per test produced when equal sample sizes were selected from the control population and the experimental populations. Under this sampling condition, the number of Type I error rates per test recorded for each of the four non-nil null hypotheses ranged from .050 to .051. None of these figures differed from the number of Type I error rates per test recorded for the nil null hypothesis, which was .050, by more than .001. All of the values were close to or equal to the established alpha level of .05.

The Type I error rates per test listed under the column entitled "Combination 2" in Table 2, were recorded for the testing situation in which the sample sizes were unequal, i.e., 15 and 25 scores were randomly sampled from the control population and each of the experimental populations, respectively. The number of Type I error rates produced for the four non-nil null hypotheses ranged from .049 to .054, which differed from the rate of .050 recorded for the nil null by no more than .004. Again, all of the Type I error rates per test were close to the established alpha level of .05.

The figures listed under the column entitled "Combination 3" in Table 2 reflected the Type I error rates per test produced by the testing situation in which the sample sizes were again unequal. Under this condition, however, 25 were randomly sampled from the control population and 15 scores were randomly selected from each of the experimental populations. The number of Type I error rates produced for the four non-nil null hypotheses, which ranged from .044 to .053, differed from the .051 figure recorded for the nil null hypothesis by no more than .007. Once again, all of the Type I error rate values were close to or equal to the established alpha level of .05.

Unequal variances for normally distributed populations. The results listed in Table 3 were produced under the condition where the variance of the control population, C2, was one half the size of the variances for the experimental populations, E6 through E10. In Table 3, the numbers listed under the column entitled "Combination 4" indicated the number of Type I error rates per test produced when equal sample sizes were selected from the control population and the experimental populations. Under this sampling condition, the number of Type I error rates per test recorded for each of the four non-nil null hypotheses ranged from .047 to .056. None of these figures differed from the number of Type I error rates per test recorded for the nil null hypothesis, which was .050, by more than .006. All of the values for the nil null hypotheses and the non-nil null hypotheses were close to or equal to the established alpha level of .05.

The Type I error rates per test listed under the column entitled "Combination 5" in Table 3, were recorded for the testing situation in which the sample sizes were unequal, i.e., 15 and 25 scores were randomly sampled from the control population and each of the experimental populations, respectively. The number of Type I error rates produced for the four non-nil null

hypotheses ranged from .046 to .054, which differed from the rate of .056 recorded for the nil null by no more than .012. It should be noted that each of the Type I error rates for the four non-nil null hypotheses were within .006 of the established alpha level of .05.

The figures listed under the column entitled "Combination 6" in Table 3 were the Type I error rates per test produced by the testing situation in which the sample sizes were again unequal. Under this condition, however, 25 and 15 scores were randomly sampled from the control population and each of the experimental populations, respectively. The number of Type I error rates produced for the four non-nil null hypotheses, which ranged from .047 to .052, differed from the figure of .051 recorded for the nil null hypothesis by no more than .004. Once again, all the Type I error rates were close to or equal to the established alpha level of .05.

Similar to the results listed in Table 3, the results listed in Table 4 were produced under the condition where the variance of the control population, C3, and the experimental populations, E2 through E5, differed. For the results contained in Table 4, however, the variance for C3 was twice the size of the variances for the experimental populations E2 through E5. In Table 4, the numbers listed under the column entitled "Combination 7" revealed the number of Type I error rates per test produced when equal sample sizes were selected from the control population and the experimental populations. The number of Type I error rates per test recorded for each of the four non-nil null hypotheses ranged from .047 to .054. None of these figures differed from the number of Type I error rates per test recorded for the nil null hypothesis, which was .052, by more than .005. All the values recorded for the four non-nil null hypotheses were close to the established alpha level of .05.

The Type I error rates per test listed under the column entitled "Combination 8" in Table 4, were recorded for unequal sample sizes. Fifteen scores and 25 scores were randomly sampled from the control population and each of the experimental populations, respectively. The number of Type I error rates produced for the four non-nil null hypotheses ranged from .050 to .054, which differed from the rate of .049 recorded for the nil null by no more than .005. Each of the Type I error rates for the four non-nil null hypotheses were within .004 of the established alpha level of .05.

The Type I error rates listed under the column entitled "Combination 9" in Table 4 were also generated under the condition of unequal sample sizes. Under this condition, however, 25 scores and 15 scores were randomly sampled from the control population and each of the experimental populations, respectively. The number of Type I error rates produced for the four non-nil null hypotheses ranged from .047 to .054. These figures differed by no more than .012 from the .051 figure generated for the nil null hypothesis. All the Type I error rates were within a value of .004 of the established alpha level of .05.

Non-Normally distributed populations. Non-normally distributed populations were used to generate the results contained in Table 5. The Type I error rated listed under the column entitled "Combination 10" were produced from equal sample sizes. Under this sampling condition, the number of Type I error rates per test recorded for the non-nil null hypothesis was .051, which differed only slightly from the rate of .044 produced for the nil null hypothesis. The .051 rate produced for the non-nil null hypothesis was very close to the established alpha level of .05.

The Type I error rates per test listed under the column entitled "Combination 11" in Table 5, were recorded for unequal sample sizes. Fifteen scores and 25 scores were randomly sampled from the control population and each of the experimental populations, respectively. The number of Type I error rates produced for the non-nil null hypothesis was .047, which again differed only slightly from the .054 recorded for the nil null hypothesis. The .047 rate generated for the nil null hypothesis was close to the established alpha level of .05.

The figures listed under the column entitled "Combination 12" represented the Type I error rates per test produced when 25 scores were randomly sampled from the control population and 15 scores were randomly selected from each of the experimental populations. The error rate produced for the non-nil null hypothesis was .056. This figure was close to both the rate recorded for the nil null hypothesis, which was .049, and the established alpha level of .05.

Summary of results. The Type I error rates generated for the nil-null hypotheses and the non-nil null hypotheses were very similar. The Type I error rates per test recorded for the non-nil null hypotheses were also close to the established alpha level. These results were produced for each of the following conditions: (a) normally distributed populations, (b) non-normally distributed populations, (c) equal and unequal sample sizes, (d) equal and unequal population variances, and (e) various sizes of the values incorporated into the nil null hypotheses.

Summary and Implications

The results presented in this study suggest the Type I error rates of independent t tests used to test nil null hypotheses and non-nil null hypotheses do not differ to any meaningful degree. Even though the generalizability of these results are limited by the finite combination of population parameters and sample sizes examined in this study, they should provide additional

encouragement for researchers to employ independent t tests to statistically test non-nil null hypotheses that incorporate values deemed to be practically significant by researchers and practitioners.

We believe the use of such hypotheses will lead to better quality research for the following three reasons.

1. This type of non-nil null hypotheses reflects the viewpoint that the concepts of practical and statistical significance are both essential components of the evaluation process.
2. The use of non-nil null hypotheses requires researchers and practitioners to identify and justify the level by which the two groups must differ in order for the difference to be viewed as clinically or educationally important. Thus, researchers must reflect on the relevance of their research questions, i.e., researchers should not approach the research process in a mechanical manner.
3. The practical significance level, which is incorporated into the non-nil null hypothesis, may be most beneficial if it is not be based on statistical concerns. That is, in the process of identifying the practical significance level, researchers can, and we believe should, consider societal concerns and cost versus benefit comparisons.

We believe an increased use in non-nil null hypotheses that incorporate practical significance levels, will require further developments in the methods used to identify practical levels of significance. Our experience has shown that the identification of practical or clinically significant values incorporated into the non-nil null hypotheses, is a difficult task for researchers and practitioners. Discussions of various methods and philosophical positions regarding the

identification of practical significance values would assist researchers and practitioners who undertake such a task. Such assistance would encourage researchers to pose research questions that include practical significance levels. These types of questions would require the use of non-nil null hypotheses. We believe such research practices would enhance the quality of research being conducted.

In addition to developing various methods and philosophical positions regarding the identification of practical significance values, we believe it is important for researchers to undertake future studies that investigate possible limitations of t tests and F tests used in conjunction with non-nil null hypotheses. If certain research designs reveal limitations of t and F tests, researchers may need to consider using other procedures for testing non-nil null hypotheses in those types of designs. Such testing procedures may include randomization tests.

References

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1994). The earth is round ($p < .05$). American Psychologist, 49, 997-1003.

Fraas, J. W., & Newman, I. (2000, October). Testing for statistical and practical significance: A suggested technique using a randomization test. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL.

Fraas, J. W., & Newman, I. (2001, February). Testing Non-Nil Null Hypotheses Using Resampling Stats Add-In for Microsoft Excel. Paper presented at the annual meeting of the Eastern Educational Research Association, Hilton Head Island, SC.

Huberty, C. J. (1993). Historical origins of statistical testing practices: The treatment of Fisher versus Neyman-Pearson views in textbooks. The Journal of Experimental Education, 61 (4), 317-333.

Microsoft Excel Professional 2000 [Computer software]. (2000). Seattle, WA: Microsoft, Inc.

Newman, I., Deitchman, R., Burkholder, J., Sanders, R., and Ervin, L. (1976). Type VI error: Inconsistency between the statistical procedure and the research question. Multiple Linear Regression Viewpoints, 6(4), 1-19.

Robinson, D. H. & Levin, J. R. (1997). Reflections on statistical and substantive significance, with a slice of replication. Educational Researcher, 26 (5), 21-26.

Serlin, R. A. & Lapsley, D. K. (1993). Rational appraisal of psychological research and the good-enough principle. In G. Keren & C. Lewis (Eds.), A handbook for data analysis in the behavioral sciences: Methodological issues (pp. 199-228). Hillsdale, NJ: Erlbaum.

Serlin, R. C. & Lapsley, D. K. (1985). Rationality in psychological research: The good-enough principle. American Psychologist 40 (1), 73-83.

Shaver, J. P. (1993). What statistical significance testing is, and what it is not. Journal of Experimental Education, 61 (4), 293-316.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. Educational Researcher, 25 (2), 26-30.

Thompson, B. (1999a). If statistical significance tests are broken/misused, what practices should supplement or replace them? Theory and Psychology, 9 (2), 165-181.

Thompson, B. (1999b). Statistical significance tests, effect size reporting and the vain pursuit of pseudo-objectivity. Theory and Psychology, 9 (2), 191-196.

Thompson, B. (1999c). Journal editorial policies regarding statistical significance tests: Heat is to fire as p is to importance. Educational Psychology Review, 11 (2), 157-169.

Table 1

Mean and Variances of the Control and Experimental Populations

Populations	Parameters	
Normally Distributed	Mean	Varinance
C1	25.0	2.0
C2	25.0	4.0
E1	25.0	2.0
E2	25.3	2.0
E3	25.7	2.0
E4	26.1	2.0
E5	30.0	2.0
E6	25.0	4.0
E7	25.3	4.0
E8	25.7	4.0
E9	26.1	4.0
E10	30.0	4.0
Non-Normally Distributed		
C3	25.0	25.0
E11	25.0	25.0
E12	30.0	30.0

Table 2

Type I Error Rates Per Test for Various Sample Sizes Randomly Selected From Normally Distributed Populations: Variances of the Control and Experimental Populations Are Equal [a]

	Various Parameter and Sample Size Combinations [b,c]		
Null Hypotheses	Combination 1 (n_{C1}=$n_{E1\text{-}E5}$=25)	Combination 2 (n_{C1}=15; $n_{E1\text{-}E5}$=25)	Combination 3 (n_{C1}=25; $n_{E1\text{-}E5}$=15)
Nil Null Hypothesis			
$1H_0$: $\mu_{C1} = \mu_{E1}$	.050	.050	.051
Non-Nil Null Hypotheses			
$1H_0$: $\mu_{C1} = \mu_{E2}$ - .3	.050	.054	.045
$2H_0$: $\mu_{C1} = \mu_{E3}$ - .7	.051	.054	.050
$3H_0$: $\mu_{C1} = \mu_{E4}$ - 1.1	.051	.049	.053
$4H_0$: $\mu_{C1} = \mu_{E5}$ - 5.0	.050	.053	.044

[a] Results are based on 5,000 replications with the alpha level set equal to .05 for each test. Each figure represents the proportion of Type I errors committed.

[b] The variances of the populations are $\sigma^2_{C1} = \sigma^2_{E1\text{-}E5} = 2$.

[c] The population means are $\mu_{C1} = 25.0$, $\mu_{E1} = 25.0$, $\mu_{E2} = 25.3$, $\mu_{E3} = 25.7$, $\mu_{E4} = 26.1$, and, $\mu_{E5} = 30.0$.

Table 3

Type I Error Rates Per Test for Various Sample Sizes Randomly Selected From Normally Distributed Populations: Variance of the Control Population Equal Is Equal to 2 and the Variances of the Experimental Groups Equal Are Equal to 4 [a]

	Various Parameter and Sample Size Combinations [b c]		
Null Hypotheses	Combination 4 (n_{C1}=$n_{E6\text{-}E10}$=25)	Combination 5 (n_{C1}=15; $n_{E6\text{-}E10}$=25)	Combination 6 (n_{C1}=25; $n_{E6\text{-}E10}$=15)
Nil Null Hypothesis			
$2H_0$: $\mu_{C1} = \mu_{E6}$	.050	.056	.051
Non-Nil Null Hypotheses			
$5H_0$: $\mu_{C1} = \mu_{E7} - .3$	.047	.051	.050
$6H_0$: $\mu_{C1} = \mu_{E8} - .7$	.052	.054	.052
$7H_0$: $\mu_{C1} = \mu_{E9} - 1.1$	.050	.046	.050
$8H_0$: $\mu_{C1} = \mu_{E10} - 5.0$	.056	.048	.047

[a] Results are based on 5,000 replications with the alpha level set equal to .05 for each test. Each figure represents the proportion of Type I errors committed.

[b] Variances of the populations are $\sigma^2_{C1} = 2$ and $\sigma^2_{E6\text{-}E10} = 4$.

[c] The population means are $\mu_{C1} = 25.0$, $\mu_{E6} = 25.0$, $\mu_{E7} = 25.3$, $\mu_{E8} = 25.7$, $\mu_{E9} = 26.1$, and, $\mu_{E10} = 30.0$.

Table 4

Type I Error Rates Per Test for Various Sample Sizes Randomly Selected From Normally Distributed Populations: Variance of the Control Population Equal Is Equal to 4 and the Variances of the Experimental Groups Equal Are Equal to 2 [a]

	Various Parameter and Sample Size Combinations [b,c]		
Null Hypotheses	Combination 7 (n_{C2}=$n_{E1\text{-}E5}$=25)	Combination 8 (n_{C2}=15; $n_{E1\text{-}E5}$=25)	Combination 9 (n_{C2}=25; $n_{E1\text{-}E5}$=15)
Nil Null Hypothesis			
$3H_0$: $\mu_{C2} = \mu_{E1}$	.052	.049	.059
Non-Nil Null Hypotheses			
$9H_0$: $\mu_{C2} = \mu_{E2}$ - .3	.051	.050	.048
$10H_0$: $\mu_{C2} = \mu_{E3}$ - .7	.054	.053	.051
$11H_0$: $\mu_{C2} = \mu_{E4}$ - 1.1	.051	.054	.054
$12H_0$: $\mu_{C2} = \mu_{E5}$ - 5.0	.047	.053	.047

[a] Results are based on 5,000 replications with the alpha level set equal to .05 for each test. Each figure represents the proportion of Type I errors committed.

[b] Variances of the populations are $\sigma^2_{C2} = 4$ and $\sigma^2_{E1\text{-}E5} = 2$.

[c] The population means are $\mu_{C2} = 25.0$, $\mu_{E1} = 25.0$, $\mu_{E2} = 25.3$, $\mu_{E3} = 25.7$, $\mu_{E4} = 26.1$, and, $\mu_{E5} = 30.0$.

Table 5

Type I Error Rates Per Test for Unequal Variances From

Poisson Distributed Populations with Equal and Unequal Group Sample Sizes [a]

	Various Parameter and Sample Size Combinations [b c]		
Null Hypotheses	Combination 10 (n_{C3}=25; $n_{E11\text{-}E12}$=25)	Combination 11 (n_{C3}=15; $n_{E11\text{-}E12}$=25)	Combination 12 (n_{C3}=25; $n_{E11\text{-}E12}$= 15)
Nil Null Hypothesis			
$4H_0$: $\mu_{C3} = \mu_{E11}$	.044	.054	.049
Non-Nil Null Hypothesis			
$13H_0$: $\mu_{C3} = \mu_{E12} - 5.0$	.051	.047	.056

[a] Results are based on 5,000 replications with the alpha level set equal to .05 for each test. Each figure represents the proportion of Type I errors committed.

[b] The variances of the populations are σ^2_{C3}=25 and $\sigma^2_{E11\text{-}E12}$=30.

[c] The population means are μ_{C3} = 25.0, μ_{E11} = 30.0, and μ_{E12} = 30.0.